



#18

Docket No. AUS000168US1

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: **Shrader et al.**
Serial No.: **09/578,751**
Filed: **May 25, 2000**
For: **Client-Side Pricing Agent for
Collecting and Managing Product
Price Information Over the Internet**

§
§ Group Art Unit: **3627**
§
§ Examiner: **Fischer, Andrew J.**
§
§

RECEIVED
JUN 17 2004
GROUP 3600

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

**ATTENTION: Board of Patent Appeals
and Interferences**

Certificate of Mailing Under 37 C.F.R. § 1.8(a)

I hereby certify this correspondence is being deposited with the United States Postal Service as First Class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on June 7th, 2004.

By:

Amelia C. Turner
Amelia C. Turner

06/14/2004 JADD01 00000117 090447 09578751
02 FC:1402 330.00 DA

APPELLANT'S BRIEF (37 C.F.R. 1.192)

This brief is in furtherance of the Notice of Appeal, filed in this case on February 5, 2004.

The fees required under § 1.17(c), and any required petition for extension of time for filing this brief and fees therefore, are dealt with in the accompanying TRANSMITTAL OF APPEAL BRIEF.

This brief is transmitted in triplicate. (37 C.F.R. 1.192(a))

REAL PARTIES IN INTEREST

The real party in interest in this appeal is the following party: IBM Corporation

RELATED APPEALS AND INTERFERENCES

With respect to other appeals or interferences that will directly affect, or be directly affected by, or have a bearing on the Board's decision in the pending appeal, there are no such appeals or interferences.

STATUS OF CLAIMS

A. TOTAL NUMBER OF CLAIMS IN APPLICATION

Claims in the application are: 1-14

B. STATUS OF ALL THE CLAIMS IN APPLICATION

1. Claims canceled: 15-20
2. Claims withdrawn from consideration but not canceled: NONE
3. Claims pending: 1-14
4. Claims allowed: NONE
5. Claims rejected: 1-14

C. CLAIMS ON APPEAL

The claims on appeal are: 1-14

STATUS OF AMENDMENTS

An amendment after final rejection is submitted herewith. By this amendment, claim 15 is canceled to simplify and materially reduce issues for appeal. Because the amendment merely cancels claim 15 without any other amendments, Appellants submit that the amendment does not require further search or consideration and should be entered.

SUMMARY OF INVENTION

A client-side application enables a user to collect and manage product price information retrieved from various servers in a computer network such as the Internet. See specification, page 11, line 16, to page 12, line 4. The pricing agent is used to generate a profile at a client computer, wherein the profile identifies a given site URL, an item to be queried, and a scan interval. See specification, page 21, line 9, to page 23, line 15. The pricing agent retrieves data from the given site URL on a periodic basis as defined by the scan interval and parses the retrieved data according to a site template at the client computer to generate a data record comprising an item name and an associated price value. See specification, page 23, line 16, to page 29, line 11.

ISSUES

The issues on appeal are as follows:

Whether claims 1-14 are unpatentable as being obvious over the Streetprices Web site as known by others and in public use before the date of invention.

GROUPING OF CLAIMS

The claims on appeal do not stand or fall in a single group, but are grouped into the following groups for the reasons set forth in the Argument section below:

Claims 1-7 form group A. Claim 8 forms group B. Claim 9 forms group C. Claim 10 forms group D. Claim 11 forms group E. Claims 12-14 form group F.

ARGUMENT

The Office Action rejects claims 1-14 under 35 U.S.C. § 103 as being unpatentable over Streetprices 1/99. This rejection is respectfully traversed.

I. Claims 1-14 Must Be Considered on Their Own Merits

With respect to claims 1-14, the Office Action relies on remarks made in the Response to Restriction Requirement as allegedly admitting that all of claims 1-14 are not patentably distinct from claims 15 and 16. The Final Office Action rejected claim 15 under 35 U.S.C. § 102(b) as being anticipated by the streetprices.com Web site. The rejection of claims 1-14 is based solely on the alleged admission. Appellants note that claims 15 and 16 are now canceled, leaving only one inventive group defined in the Restriction Requirement. Therefore, any alleged admission is now moot. As such, claims 1-14 must be considered on their own merits, as they should have been throughout prosecution.

Furthermore, an alleged admission that two inventions are not patentably distinct is in no way an admission that individual claims **within an inventive group** do not have further patentable features. Appellants do not attempt to separately argue the patentability of both inventive groups. The Response to Restriction Requirement argued that claims 15 and 16 from Invention II are similar in scope to independent claim 1 of Invention I. It follows that claims 15 and 16 of Invention II are less limiting than dependent claims 2-9 of Invention I. Appellants have never argued that each of claims 1-14, individually, are not patentably distinct from claim 15, because it was only necessary to show that claims 15 and 16 were not patentably distinct from claim 1 to overcome the restriction requirement. Claims 2-14 may be within Invention I and still include further patentable features.

The Examiner bears the burden of establishing a *prima facie* case of obviousness based on the prior art when rejecting claims under 35 U.S.C. § 103. *In re Fritch*, 972 F.2d 1260, 23 U.S.P.Q.2d 1780 (Fed. Cir. 1992). The Examiner may not use the existence of a less limiting inventive group to escape the burden of establishing a *prima facie* case of obviousness for the more limiting group. Individual claims within an inventive group may recite further combinations of features not shown or suggested by the prior art. By specifically addressing only claim 15, the Examiner has only addressed the corresponding limitations in claims 1 and 10 in Invention I. The Examiner apparently agrees that independent claim 10 has different scope from independent claim 1, because the Examiner did not issue a duplicate claim rejection. The Examiner also apparently agrees that dependent claims 2-9 and 11-14 further limit the claims on which they depend, because the Examiner did not issue a rejection under 35 U.S.C. § 112, fourth

paragraph. Therefore, Appellants assert that the Examiner was aware of his burden all along and that claims 1-14 must now be considered on their own merits, as they should have been throughout prosecution.

II. The Rejection of Claims 1-14 Falls with the Rejection of Claim 15

With respect to claims 1-14, the Office Action relies on remarks made in the Response to Restriction Requirement as allegedly admitting that all of claims 1-14 are not patentably distinct from claims 15 and 16. The Final Office Action rejected claim 15 under 35 U.S.C. § 102(b) as being anticipated by the streetprices.com Web site. The rejection of claims 1-14 is based solely on the alleged admission. No one of claims 1-14 is specifically or individually addressed in the Final Office Action. Therefore, if the rejection of claim 15 falls, then so does the rejection of claims 1-14. For this reason, the rejection of claim 15, now canceled, will be addressed below.

III. The Prior Art Does Not Anticipate the Claimed Invention (Groups A-F)

The Office Action rejects claim 15 under 35 U.S.C. § 102 as being anticipated by 'StreetPrices.com' website printout of January 25, 1999 ("Streetprices 1/99"). This rejection is respectfully traversed.

With respect to claim 15, the Office Action states:

Streetprices 1/99 discloses generating a set of product profiles (see e.g. 'Computer', 'Flowers' and 'Toys'), each identifying a given site URL (inherent); a list of one or more included items to be queried (inherent), a scan interval (inherent in the graphs) and a site template (inherent in the search engine); periodically retrieving data from the given site URL (inherent in the x axis of the graph) according to the scan interval; parsing the data retrieved to generate a data record (inherent since it discloses the price over time); storing the data records (inherent); and the product profiles include one or more triggers (a threshold price which sends an email when a price drops below a particular level).

Claims 15 and 16 are alternatively rejected under 35 U.S.C. 102(b) based upon a public use or sale of the invention. In this case, evidence to support the public use is the StreetPrices.com website. See the additional cited StreetPrices.com printout noted below and the Streetprices 1/99 noted above.


Office Action, dated July 29, 2003. Applicant respectfully disagrees. The streetprices.com Web site does not explicitly or inherently teach the features of the presently claimed invention and, thus, does not anticipate the claimed invention.

A. **Lexicography**

The Examiner concludes that Appellants have decided not to be their own lexicographer by indicating and defining claim limitations to have meanings other than their ordinary and accustomed meanings. The Examiner then adopts definitions, under the broadest reasonable interpretation standard, in the claim interpretations for the purposes of applying prior art. The definitions are as follows:

- Server:** 2. On the Internet or other network, a computer or program that responds to commands from a client. Taken from *Computer Dictionary*, 3rd Edition, Microsoft Press, Redmond, WA, 1997.
- Client:** 3. On a local area network or Internet, a computer that accesses shared network resources provided by another computer (called a *server*). *Id.*
- Computer:** Any machine that does three things: accepts structured input, processes it according to prescribed rules, and produces the results as output. *Id.*
- Data:** Plural of the Latin *datum*, meaning an item of information. In practice, data is often used for the singular as well as plural form of the noun. *Id.*
- Web site:** A group of related HTML documents and associated files, scripts, and databases that is served up by an HTTP server on the World Wide Web. The HTML documents in a Web site generally cover one or more related topics and are interconnected through hyperlinks. Most Web sites have a home page and their starting point, which frequently functions as a table of contents for the site. Many large organizations, such as corporations, will have one or more HTTP servers dedicated to a single Web site. However, an HTTP server can also serve several small Web sites, such as those owned by individuals. Users need a Web browser and an Internet connection to access a Web site. *Id.*
- HTTP Server:** 1. Server software that uses HTTP to serve up HTML documents and any associated files and scripts when requested by a client, such as a Web browser. The connection between client and server is usually broken after the requested document or file has been served. HTTP servers are used on Web and Intranet sites. *Also called* Web server... 2. Any machine on which an HTTP server program is running. *Id.*
- Tag:** 1. In programming, one or more characters containing information about a file, record type, or other structure. 2. In certain types of data files, a key or an address that identifies a record and its storage location in another file... 3. In markup languages such as SGML and HTML, a code that identifies an element in a document, such as heading or a paragraph, for the purposes of formatting, indexing, and linking information in the document. *Id.*

Appellants acknowledge that definitions found in the present specification for specific terms in the claims are not being relied upon at this time. Nevertheless, Appellants submit that Appellants **always** have the right to be their own lexicographers and reserve this right for terms that may be introduced into the claims later in prosecution. Appellants accept the above



definitions for the purposes of this appeal; however, Appellants submit that the above definitions are not the only definitions that may be known to a person of ordinary skill in the art. Other technical dictionaries and other resources exist and may be relied upon. As further evidence that other definitions may exist for these terms, note that the Examiner provides multiple definitions from the same source for particular terms and, in some cases, has chosen a single definition when more than one is given. Nevertheless, the above definitions are not unreasonable when interpreted in light of the present specification and are adequate for analyzing the prior art with respect to the presently claimed invention.

B. The prior Art Does Not Teach or Suggest a Method or Computer Program Product Where the Functions are Performed in a Client Computer (Groups A-F)

The streetprices.com Web site is clearly an HTTP server or Web site that is embodied on a server. From the teachings of the streetprices.com references, a person of ordinary skill in the art will recognize that the streetprices.com Web site or HTTP server receives a search query identifying a given product, obtains prices for the given product from selected e-commerce Web sites, and generates a return HTML document that presents the identified prices for the given product. Thus, the streetprices.com Web server performs the functions of the Web site, generates Web pages, and returns these pages to the client. As a person of ordinary skill in the art would instantly recognize, a browser client application on a client computer may request HTML Web pages from the streetprices.com Web site. The only functions being performed at the client computer are requesting HTML documents, rendering HTML documents, displaying HTML documents, and, responsive to a user selecting a link in an HTML document, requesting further HTML documents.

In contradistinction, the present invention provides a method and computer program product at a client computer for collecting product data. Claim 15 (now canceled) recited:

15. A computer program product having a set of instructions executable by a processor for use **in a client computer** for collecting product data, the computer program product comprising:
 - instructions for generating a set of product profiles each identifying a given site URL, a list of one or more included items to be queried, a scan interval, and a site template;
 - instructions for periodically retrieving data for a given product profile from the given site URL according to the scan interval; and

instructions for parsing the data retrieved from the given site URL according to the site template to generate a data record for each included item comprising at least an item name and an associated price value. [emphasis added]

Claim 1 recites:

1. A method **in a client computer** for collecting product data, comprising the computer implemented steps of:
generating a profile identifying a given site URL, an item to be queried, and a scan interval;
on a periodic basis as defined by the scan interval, retrieving data from the given site URL; and
parsing the retrieved data according to a site template to generate a data record comprising an item name and an associated price value. [emphasis added]

Claim 10 recites:

10. A method, **in a client computer**, of collecting product data, comprising the steps of:
generating a set of product profiles, each identifying a given site URL, a list of one or more included items to be queried, a scan interval, and a site template;
for a given product profile, periodically retrieving data from the given site URL according to the scan interval;
parsing the data retrieved from the given site URL according to the site template to generate a data record for each included item comprising an item name, an associated price value, and a secondary source; and
storing the data records. [emphasis added]

Implementing the functions of the claimed invention in a client computer presents several advantages. When a user generates a set of product profiles at a client computer, the user has much more control over the collection of product data. By generating the set of product profiles in the client computer, the user is able to designate a given site URL to be searched. That is, the user need not rely on a Web service to work out an agreement with e-commerce Web sites; the user may use any Web site he or she desires, as long as a URL is known for the given site.

Furthermore, the user may define a scan interval in the product profiles and have the client-side agent perform the search according to the scan interval defined by the user. Thus, the user has much more control over how often the product data collection is performed. On the other hand, with the streetprices.com Web site, the user is restricted to the scan interval rigidly defined at the Web server.

Moreover, the user need not depend upon a single Web server for operation. That is, if the streetprices.com Web site is not operational, then the user is not able to collect product data. In addition, the streetprices.com Web server may perform product data collection for thousands, even millions, of users. Thus, the performance of the streetprices.com Web server may be degraded during peak usage. This single point of failure in the prior art is a disadvantage that is overcome by the present invention.

Still further, as recited in claim 10 and previously presented in claim 15, a user may define in a set of product profiles, a site template that is to be used when parsing the results from a given site. Thus, the user may control how the site is parsed by providing a specific template for the site. On the other hand, with the streetprices.com Web site, the user must rely on whatever parsing mechanisms are employed by the server software.

None of the Streetprices references teaches or suggests a computer program product "having a set of instructions executable by a processor for use in a **client computer** for collecting product data," as previously presented in claim 15. Streetprices is clearly a Web site, which is embodied and executed on a Web **server**. The Web server of the streetprices.com references clearly is not a client computer, as recited in claims 1 and 10 and previously presented in claim 15, particularly when read in light of the present specification and, more particularly, when given the definitions provided above. The Office Action proffers no analysis whatsoever as to why a Web server is somehow a client computer. In fact, the Office Action appears to make no attempt whatsoever to address this limitation. Therefore, the Office Action fails to establish a *prima facie* case of anticipation for claim 15. It follows that the rejection of claims 1-14 also falls with the rejection of claim 15 and, thus, claims 1-14 are not anticipated or rendered obvious by the streetprices.com Web site. Therefore, Appellants respectfully request that the rejection of claims 1-14 be overturned.

C. Inherency

The Office Action misapplies the concept of "inherent" anticipation. Section 102 of Title 35 deals with novelty and loss of patent rights. Under the principles of inherency, a claim is anticipated if a structure in the prior art **necessarily** functions in accordance with the limitations of a process or method claim. *In re King*, 801 F.2d 1324, 231 U.S.P.Q. 136 (Fed. Cir. 1986).

The missing claimed characteristics must be a “natural result” flowing from what is disclosed. *Continental Can Co. v. Monsanto Co.*, 948 F.2d 1264, 20 U.S.P.Q.2d 1746 (Fed. Cir. 1991). Unstated elements in a reference are inherent when they exist as a “matter of scientific fact.” *Constant v. Advanced Micro-Devices, Inc.*, 848 F.2d 1560, 7 U.S.P.Q.2d 1057 (Fed. Cir.), *cert. denied*, 488 U.S. 892 (1988) and *Hughes Aircraft Co. v. United States*, 8 U.S.P.Q.2d 1580 (Ct. Cl. 1988).

In the present case, the Examiner's assertion that these elements are present can be made only through the use of the Applicants' disclosure as a template to fill in the missing elements. The Office Action does not establish any necessity, natural result, or scientific fact to support a conclusion of inherency. The Office Action merely cites a Web site that is somewhat similar to the client-based invention and simply dismisses all of the claim limitations as inherent.

The Final Office Action states:

In light of Applicants' “robots” argument noted above, it should not be too surprising that Applicants have also misapplied inherency. The Examiner will point out some of the errors in Applicants' response and provide a brief review of inherency.

Office Action, dated November 5, 2003: Snide comment aside, Appellants submit that Appellants do not misapply inherency, because Appellants do not even attempt to apply inherency. Rather, Appellants will show that the applied prior art does not explicitly or inherently teach each and every limitation of the claimed invention.

The Examiner argues that the citation to *King* is not on point, because claim 15 is not a process or method claim. This is further evidence that the Examiner appears to recognize that claims 1-14 should have been considered on their own merits. Yet, the Examiner does not consider whether the streetprices.com Web site **necessarily** functions in accordance with the limitations of the method claims. Furthermore, the principal of the *King* citation still holds for claim 15. Appellants submit that the streetprices.com Web site does not **necessarily** function in accordance with the functional limitations of the computer program product claim previously presented in claim 15.

The Examiner recognizes that Appellants cited a “bunch” of interparty litigation cases, but only addresses *Continental Can Co. v. Monsanto Co.*, 948 F.2d 1264, 20 USPQ2d 1746 (Fed. Cir. 1991). However, even in considering this case, the Examiner addresses only a fragment of

the argument. The Examiner fails to address that a missing claimed characteristics must be a “natural result” flowing from what is disclosed.

The Examiner then goes on to define “inherent” to mean “involved in the constitution or essential character of something...,” taken from Merriam-Webster’s Collegiate Dictionary. However, the Examiner still fails to show how generating a set of product profiles in a client computer, product profiles including a given site URL, product profile including a scan interval, and storing data records in a client computer are involved in the constitution or essential character of the streetprices.com Web server.

The Examiner then goes to great lengths to support a conclusion that inherency need only be established by a preponderance of the evidence. Then, the Examiner makes the following conclusion:

Based upon the discussion above regarding inherency and the operation of Streetprices 1/99, the Examiner finds that one of ordinary skill in the art would reasonably infer that it is more likely than not that all limitations in claim 15 (including all structural and functional limitations) are disclosed either directly or inherently in Streetprices 1/99.

Appellants respectfully disagree. Despite all the legal analysis and lengthy discussion, the Examiner has presented no evidence to support a conclusion that the missing elements are explicitly taught or inherently present.

What is known from the applied references is that streetprices.com allows a user to enter a search term or click on a link for a product, collects price information from other Web sites, and generates a results page that is returned to the client. The streetprices.com references also teach that robots are used to crawl 48 e-commerce web sites. Robots, also referred to as “bots,” “software agents,” or “traveling programs,” as well as crawling, are notoriously well-known in the art. Robots are programs that automatically do some action without user intervention. In the context of a search engine, a robot usually refers to a program that mimics a browser to download Web pages automatically. See attached definition from WebmasterWorld.com. A spider is a type of robot that will download multiple pages from the same Web site. Crawling refers to the fact that the spider will look for links in the pages it downloads and “walk” or “crawl” down through a Web site. See attached definition from WebmasterWorld.com. An article entitled “All About Search Indexing Robots and Spiders” from searchtools.com is also attached for the edification of the Examiner.

There is no evidence whatsoever that a product profile is generated in a client computer, wherein the product profile includes a URL for a given site, nor does it naturally or necessarily flow from the teachings of the applied references that a product profile is generated. As an alternative possibility not taken straight from the instant claims, the robots may crawl all 48 e-commerce Web sites for every single query. There is no need for a product profile in the streetprices.com Web site. There is also no need for a given site URL to be identified in a product profile, because the streetprices.com Web site collects product information for thousands or more products for thousands or more users, rather than collecting product information for a handful of products from a given site for a single user at a client computer.

The streetprices.com Web site need not collect price data periodically. The streetprices.com Web site may collect data continually and merely record price data based on a date of last change. Alternatively, the streetprices.com Web site may look at a log listing the date of the page that was last scanned and determine whether the page needed to be scanned at all. There certainly is no evidence that a scan interval is included in a product profile generated in a client computer. As an alternative possibility not taken directly from the instant claims, the streetprices.com Web site may collect data according to a scan interval hard coded in the server software.

Furthermore, with respect to independent claim 10, the streetprices.com Web site does not inherently teach that a product profile generated in a client computer includes a site template. As an alternative possibility not gleaned from the presently claimed invention, the server software of the streetprices.com web site may strip the results of all HTML tags and use natural language understanding software to identify prices. The Web site may alternatively use proximity operators hard coded in the server software to identify prices. There is no evidence that generating a product profile in a client computer, wherein the product profile includes a site template, is involved in the constitution or essential character of the streetprices.com Web site.

For the above reasons, Appellants submit that missing elements are not inherently present in the streetprices.com Web site. Thus, the Examiner does not establish a *prima facie* case of anticipation for claim 15. It follows that claims 1-14 are neither anticipated nor rendered obvious by Streetprices 1/99.

IV. The Prior Art Does Not Render the Claimed Invention Obvious (Groups A-F)

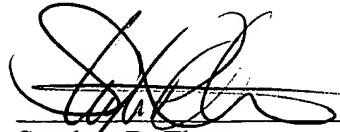
Since the rejection of claim 15 falls, so does the rejection of claims 1-14 under 35 U.S.C. § 103. Streetprices does not teach, suggest, or give any incentive to make the needed changes to reach the presently claimed invention. Streetprices actually teaches away from the presently claimed invention because it teaches a Web site embodied and executed on a Web server, as opposed to a method that is performed in a client computer, as in the presently claimed invention. In fact, the streetprices.com Web site includes advertisements, as evidenced by the applied references. Certainly, by providing a server-based Web site, the streetprices.com service wants client computers to visit the Web site so that the customer sees the advertisements. Thus, the server-based streetprices.com Web site would not lead a person of ordinary skill in the art to perform similar functions on a client, because there would be no server to control the distribution of, and exposure to, advertisements. Absent the Office Action pointing out some teaching or incentive to implement Streetprices on a client computer, one of ordinary skill in the art would not be led to modify Streetprices to reach the present invention when the reference is examined as a whole. Absent some teaching, suggestion, or incentive to modify Streetprices in this manner, the presently claimed invention can be reached only through an improper use of hindsight using the applicant's disclosure as a template to make the necessary changes to reach the claimed invention.

V. The Prior Art Does Not Teach or Suggest the Further Limitations in the Dependent Claims (Groups B, C, E, and F)

Claim 8 recites, "wherein the matching criteria is a fuzzy match"; claim 9 recites, "wherein the data record also includes secondary source information"; claim 11 recites, "wherein the given product profile also includes a list of excluded items"; and, claim 12 recites, "wherein the given product profile also includes a next scan date." The Office Action fails to address these limitations, particularly as allegedly taught or suggested prior to invention by the applicant. Therefore, the Office Action fails to establish a *prima facie* case of obviousness for these claims. Appellants submit that the streetprices.com does not teach or fairly suggest the features of claims 8, 9, 11, and 12.

VI. Conclusion

In view of the above, Appellants respectfully submit that the rejections of claims 1-14 are overcome. Accordingly, it is respectfully urged that the rejections of claims 1-14 not be sustained.



Stephen R. Tkacs
Reg. No. 46,430
Yee & Associates, P.C.
PO Box 802333
Dallas, TX 75380
(972) 367-2001

Attachments: definition of "robot" from WebmasterWorld.com; definition of "crawler" from WebmasterWorld.com; "All About Search Indexing Robots and Spiders" from SearchTools.com

APPENDIX OF CLAIMS

The text of the claims involved in the appeal reads:

1. A method in a client computer for collecting product data, comprising the computer implemented steps of:

generating a profile identifying a given site URL, an item to be queried, and a scan interval;

on a periodic basis as defined by the scan interval, retrieving data from the given site URL; and

parsing the retrieved data according to a site template to generate a data record comprising an item name and an associated price value.
2. The method as described in Claim 1 further including the step of storing the data record.
3. The method as described in Claim 1 further including the step of:

comparing the associated price data to a given threshold condition; and

taking a given action if the price data has a given relationship to the given threshold condition.
4. The method as described in Claim 3 wherein the given action is selected from the actions consisting essentially of issuing an e-mail notification, logging an event, and initiating an e-commerce transaction.

5. The method as described in Claim 1 further including the step of collecting associated price data for the item name over a given period of time to produce historical price data.
6. The method as described in Claim 1 wherein the profile includes matching criteria that must be met for the item name.
7. The method as described in Claim 7 wherein the matching criteria is an exact match.
8. The method as described in Claim 6 wherein the matching criteria is a fuzzy match.
9. The method as described in Claim 1 wherein the data record also includes secondary source information.
10. A method, in a client computer, of collecting product data, comprising the steps of:
generating a set of product profiles, each identifying a given site URL, a list of one or more included items to be queried, a scan interval, and a site template;
for a given product profile, periodically retrieving data from the given site URL according to the scan interval;
parsing the data retrieved from the given site URL according to the site template to generate a data record for each included item comprising an item name, an associated price value, and a secondary source; and
storing the data records.

11. The method as described in Claim 10 wherein the given product profile also includes a list of excluded items.

12. The method as described in Claim 10 wherein the given product profile also includes a next scan date.

13. The method as described in Claim 12 wherein the given product profile also includes a threshold expression comprising a set of one or more triggers.

14. The method as described in Claim 13 wherein the given product profile also includes a threshold action that is triggered if the threshold expression resolves to a given value.



Welcome to WebmasterWorld Guest from 66.180.32.190

[register](#), [login](#), [site search](#), [glossary](#), [subscribe](#), [help](#),
[library](#), [recent posts](#)

Subscribe to WebmasterWorld



[Home](#) / [Forums Index](#) / [Glossary](#) / [Robot](#)

Robot

A program that automatically does "some action" without user intervention. In the context of search engines, it usually refers to a program that mimics a browser to download web pages automatically. A spider is a type of robot. Some times referred to as Webbots.

See Also:

[Spider](#)

Hosted by [Westhost.com](#)

[Terms of service](#) | [Privacy Policy](#) | [Report Problem](#) | [About](#)

All trademarks and copyrights held by respective owners.

Member comments are owned by the poster.

BestBBS v3.15 (c) [WebmasterWorld.com](#) 1998-2004 *all rights reserved*



Welcome to WebmasterWorld Guest from 66.180.32.190

[register](#), [login](#), [site search](#), [glossary](#), [subscribe](#), [help](#), [library](#), [recent posts](#)

[Thanks for keeping us banner free](#)

[Home](#) / [Forums Index](#) / [Glossary](#) / [Crawler](#)

Crawler

A type of a A HREF="#spider">Spider that will download multiple pages from the same web site. Crawling refers to the fact, that the spider will look for links in the pages it downloads and then *walk* or crawl down through a web site.

See Also:

Spider

Hosted by [Westhost.com](#)

[Terms of service](#) | [Privacy Policy](#) | [Report Problem](#) | [About](#)

All trademarks and copyrights held by respective owners.

Member comments are owned by the poster.

BestBBS v3.15 (c) **WebmasterWorld.com** 1998-2004 *all rights reserved*

[Home](#) [Guide](#) [Tools List](#) [News](#) [Background](#) [Search](#) [About Us](#)

SearchTools.com

All About Search Indexing Robots and Spiders

Many search engines use programs called *robots* to gather web pages for indexing. These programs are not limited to a pre-defined list of web pages, they can follow links on pages they find, which makes them a form of *intelligent agent*. The process of following links is called *spidering*, *wandering*, or *gathering*.

Controlling Robot Indexing

Robot spiders cannot index unlinked files, so they will ignore all the miscellaneous files you may have in your web server directory. Webmasters can control which directories the robots should index by editing the [robots.txt](#) file, and web page creators can control robot indexing behavior using the [Robots META](#) tag.

Following Links

Local search robot spider indexers locate files to index by following links, just like webwide search engine spiders. You specify the starting page, and these indexers will request it from the server and received it just like a browser. The indexer will store every word on the page and then follow each link on that page, indexing the linked pages and following each link from those pages.

Link Problems

They will miss pages which have been accidentally unlinked from any of your starting points. And spiders will have problems with JavaScript links, just like webwide search engine robots.

Dynamic Elements

Robot spider indexers will receive each page exactly as a browser will receive it, with all dynamic data from CGIs, SSI (server-side includes), ASP (active server pages) and so on. This is vital to some sites, but other sites may find that the presence of these dynamic elements triggers the re-indexing process, although none of the actual text of the page has been changed.

Most site search engines can handle dynamic URLs (including question marks ? and other punctuation). However, most webwide search engines will not index these pages: for help building plain URLs, see our page on [Generating Simple URLs](#).

Server Load

Because they use HTTP, robot spider indexers can be slower than local file indexers, and can put more pressure on your web server, as they ask for each page.

Updating Indexes

To update the index, some robot spider will query the web server about the status of each linked page by asking for the HTTP header using a "HEAD" request (the usual request for an HTML page is a "GET"). For HEAD requests, the server may be able to send the page header information from an internal cache, without opening and reading the entire file, and so the interaction may be much more efficient. Then the indexer compares the modified date from the header with its own date for the last time the index was updated. If the page has not been changed, it doesn't have to

update the index. If it has been changed, or if it is new and has not yet been indexed, the robot spider will then send a GET request for the entire page, and store every word. An alternate solution is for robot spiders to send an "If-Modified-Since" request: this HTTP/1.1 header option allows the web server to send back a code if the page has not changed, and the entire page if it has changed.

Duplicate Files

Robots must contain special code to check for duplicate pages, due to server mirroring, alternate default page names, mistakes in relative file naming (. / instead of . . /, for example), and so on. Some search indexers have powerful algorithms to identify these duplicates and only store and search one copy.

For more information, see the [SearchTools Indexing Robot Checklist](#).

Robot Information Sites

Web Robots Pages

Compilation of information about robots by the author of the [Robots Exclusion Protocol](#), Martijn Koster. Slightly obsolete as of mid-1999. Includes the helpful [Web Robots FAQ](#) for Web users, authors and Robot implementors, [Web Robots Database](#) of User-Agent names and contact information, the [Guidelines for Robot Writers](#) and more.

Note that this is now hosted at the new site run by Martijn at www.robotstxt.org.

W3C HTML 4.0 Specification, Appendix B, Notes on helping search engines index your Web site

Standard information about data for indexing, robots.txt and META Robots tag.

SearchTools Robots Pages

robots.txt Page

Describes the robots.txt file format and implications for search indexing robots.

META Robots Tag Page

Describes the META Robots tag contents and implications for search indexing robots.

Indexing Robot Checklist

A list of important items for those creating robots for search indexing.

Generating Simple URLs for Search Engine Robots

All about URL Rewriting to convert dynamic URLs to simple path-based URLs.

List of Robot Source Code

Links to free and commercial source code for robot indexing spiders

List of Robot Development Consultants

Consultants who can provide services in this area.

List of Books and Articles about Search Indexing Robots

Links to writings on about Robots, Spiders and Crawlers.

Listings of Robot "User Agent" Names

Web Robots Database

List at robotstxt.org, may not be current.

SearchEngineWatch SpiderSpotting Chart

Displays User Agent and host names for webwide search engine robot spiders.

Agents and Robots List - WebReference.com

lightly annotated listing of agent and robot software

Search Engine Robots

Lists of search engines, agent names and their information links, updated fairly frequently.

Other Good SitesBotSpot

Contains many listings of robots on the Web, links, articles and bibliographies, but is not well organized and is rarely updated.

SearchEngineWorld tests of Robots.txt

Common problems found when testing their robots.txt validator.

Robots Mailing List - for writers of web robots

To subscribe, send a message to listserv@mccmedia.com with the words **subscribe robots** (your name) in the message body.

For mailing list help, see the [Listserv help message](#).

To view earlier messages, see the [Archive of discussions, 1995-1997](#)

Robot-related ArticlesBreadth-first crawling yields high-quality pages *Proceedings of the WWW10 Conference, May 2-5 2001, Hong Kong, by Mark Najork and Janet L. Wiener*

When crawling large sites, you may not be able to get every page right away. The question becomes, should you follow links in the order encountered, or try to create a priority list? This paper describes an experiment comparing results using different crawls: *breadth-first*: order discovered; *backlink*: highest known links crawled first; *PageRank*: highest PageRank quality crawled first (as defined by Google founders Page and Brin); *random*: next page selected at random. This experiment was performed on 328 million pages on 7 million host servers, and found that the first few million pages encountered were the best, because "the most important pages have many links to them from numerous hosts, and those links will be found early, regardless of on which host or page the crawl originates." Therefore, there is little need to compute PageRanks for all available pages to crawl those first.

Robot TestingSearchTools Robots Testing

A testbed which provides cases for common robot problems, including robots.txt directives, duplicate files, odd relative links, JavaScript-generated text and links, and more.

Last Modified: 2002-06-18

[Home](#) [Guide](#) [Tools Listing](#) [News](#) [Background](#) [Search](#) [About Us](#)

SearchTools.com

Copyright © 1999-2003 [Search Tools Consulting](#)